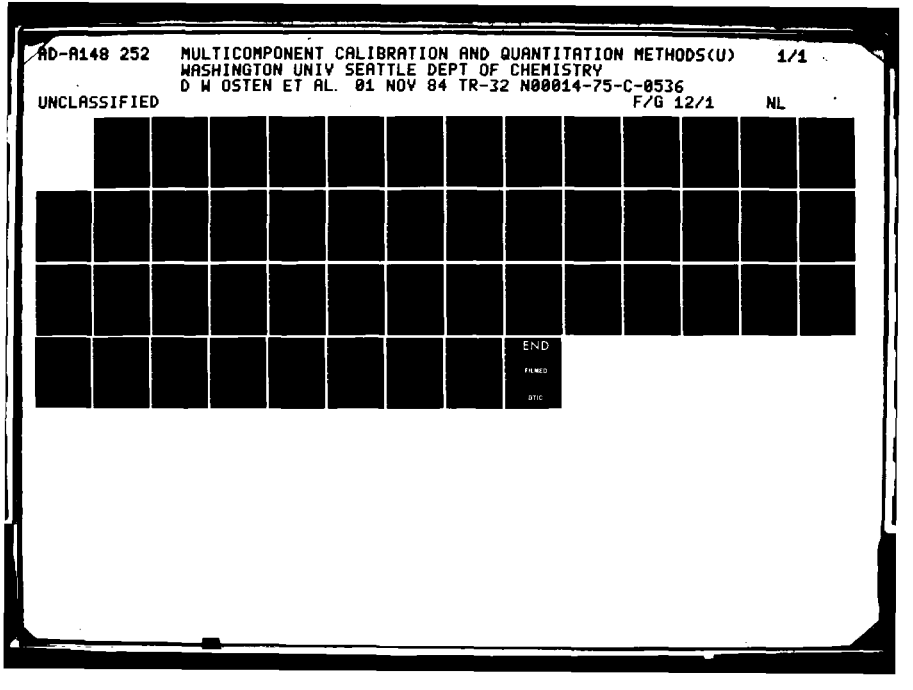
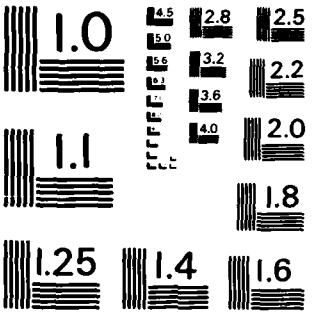

AD-A148 252
MULTICOMPONENT CALIBRATION AND QUANTITATION METHODS(U)
1/1
WASHINGTON UNIV SEATTLE DEPT OF CHEMISTRY
D W OSTEN ET AL. 01 NOV 84 TR-32 N00014-75-C-0536
UNCLASSIFIED
F/G 12/1
NL
END
FILMED
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A148 252

(12)

OFFICE OF NAVAL RESEARCH

Contract N00014-75-C-0536

Task No. NR 051-565

TECHNICAL REPORT NO. 32

Multicomponent Calibration and Quantitation Methods

by

D. W. Osten and B. R. Kowalski

Prepared for Publication

in

ASTM Standard Technical Publications

University of Washington
Department of Chemistry BG-10
Seattle, Washington 98195

DTIC
ELECTE
DEC 4 1984
A

November 1, 1984

Reproduction in whole or in part is permitted for
any purpose of the United States Government

This document has been approved for public release
and sale; its distribution is unlimited

DTIC FILE COPY

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER TR32	2. GOVT ACCESSION NO. AD-A148252	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Multicomponent Calibration and Quantitation Methods		5. TYPE OF REPORT & PERIOD COVERED Technical Report - Interim 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) D. W. Osten and B. R. Kowalski		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0536
9. PERFORMING ORGANIZATION NAME AND ADDRESS Laboratory for Chemometrics Department of Chemistry BG-10 University of Washington, Seattle, WA 98195		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 051-565
11. CONTROLLING OFFICE NAME AND ADDRESS Materials Sciences Division Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE November 1, 1984 13. NUMBER OF PAGES 46
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Prepared for publication in ASTM Standard Technical Publications; accepted.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Calibration	Multivariate Analysis Methods
Chemometrics	Multivariate Calibration
Bilinear forms	Multivariate Quantitation
Multicomponent Analysis	Quantitation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A systematic examination of calibration and quantitation methods available to analytical chemistry is made. First, simple linear calibration with one sensor is reviewed with an emphasis on chemical problems that can invalidate calibration models and what can be done about them. Then, shift is made to multivariate methods used for multicomponent analysis ending in a discussion of bilinear forms.

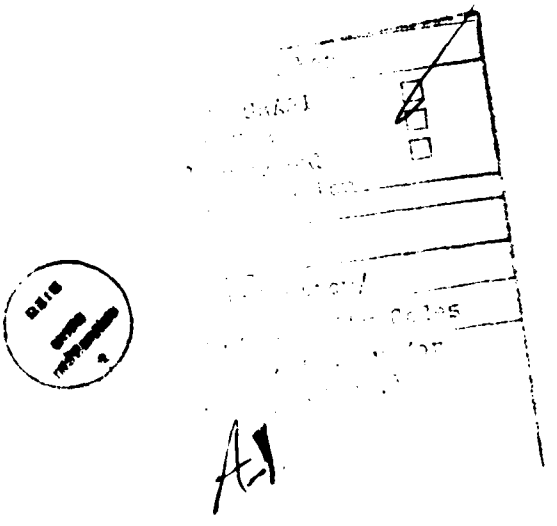
DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Abstract

A systematic examination of calibration and quantitation methods available to analytical chemistry is made. First, simple linear calibration with one sensor is reviewed with an emphasis on chemical problems that can invalidate calibration models and what can be done about them. Then, shift is made to multivariate methods used for multicomponent analysis ending in a discussion of bilinear forms.



The general problems of calibration and quantitation are well known to analytical chemists. The traditional approach to quantitative analysis has often been to find a single sensor which was specific for the desired analyte and responded in a linear manner to changes in the analyte concentration. The requirement of a fully selective sensor frequently necessitates various separation or purification steps prior to the analytical measurement. An alternative approach is to use many analytical sensors and multivariate data analysis methods. The objective of this review is to provide the reader with an overview of multivariate calibration and quantitation methods and to discuss some of the assumptions inherent in applying these approaches to analytical measurements.

Analytical Chemistry has recognized the importance of this problem by establishing a report on Chemometrics as a part of its biennial Fundamental Reviews issue. Sections of the past Chemometrics reviews entitled "Modeling and Parameter Estimation", "Calibration", and "Resolution" are of particular interest to researchers in this area (1,2).

This review is grouped into three distinct sections. A review of the single-component linear model will provide the basis for development of the more complex models, such as the multicomponent linear model. The latter is based on one-dimensional response measurements; for example, the absorption spectrum of a mixture sample. The third section will discuss the multicomponent bilinear model. This model can be used to describe two-dimensional chemical measurements such as fluorescence excitation-emission matrices, gas chromatography - mass spectrometry (GC/MS) data, liquid chromatography - UV (LC/UV) data, or spatial/spectral data as obtained from imaging in surface analysis.

Single Component Linear Model

The situation most favored by analytical chemists is when the response, r, of a single analytical sensor is a linear function of the concentration, c, of a single chemical analyte of interest.

$$r = k\,c \qquad (1)$$

In order to obtain an estimate of the analyte concentration, two general steps are required. First there is a calibration step, in which the sensitivity coefficient, k, is determined based on the analysis of one or more samples of known concentrations. The second is a quantitation step, in which the response of the unknown sample is measured and the analyte concentration is estimated from the calibration model. Implicit in using this simple linear model are a series of assumptions about the chemical system being examined: first, the response is linearly related to the analyte concentration over the concentration region of interest; second, the analytical sensor is fully specific and responds only to the analyte of interest; and third, the sensitivity coefficient does not change between the calibration and quantitation steps. If these three assumptions are obeyed for a given experimental situation, then the calibration line is obtained by measuring the response at various analyte concentrations (3).

Since the analytical sensor is fully specific, the response of a sample containing no analyte is by definition equal to zero. This implies that the calibration line must pass through the origin. In principle, measurement of a single sample of known analyte concentration is sufficient to determine the slope of the calibration line. In practice, several measurements are preferred. Even in situations where a theoretically linear relationship is known to exist, the measured experimental values will rarely be co-linear with theoretical values due to sample variance, measurement errors, and random noise.

Random Error

The method of least squares is commonly used to estimate the position of the calibration line. The mathematical formula for calculation of the least squares regression line and the confidence region around this line are well known (4,5). Several additional assumptions are required when the method of least squares is used to estimate the calibration relationship (6). The first assumption is that all the measurement error is associated with the dependent variable, the measured response. This condition requires the variance in the concentrations of the standard samples to be much smaller than the variance in the corresponding measured responses. Secondly, each measured response is drawn from a normal distribution with a mean equal to the true response for the corresponding analyte concentration. This requires that repeated measurements of the response for a single sample yield a Gaussian distribution. Lastly, the variance of the measured response must be independent of the analyte concentration, or in statistical terms there must be homogeneity of variances.

If M calibration samples have been analyzed, with each calibration sample being measured one or more times such that N total calibration measurements were made, then the calibration step requires estimating the values of two parameters; the slope, β_1, and the intercept, β_0, of the least squares line. For the single-component linear model, the least squares problem can be expressed as minimizing the sum of the squares of the residuals in the following vector equation:

$$\mathbf{r} = \beta_0 + \beta_1 \mathbf{c} + \epsilon \qquad (2)$$

where r is a column vector containing the N measured responses, c is a column

vector containing the N known analyte concentrations, and ϵ is the vector of residual errors not fitted by the model.

Shewell (7) has observed that varying the location of the calibration points will have an effect on the accuracy of the estimates obtained for the slope and the intercept of the regression line. In general, for a constant total number of calibration measurements, N, the most accurate estimate of the intercept, β_0, is obtained if N-1 measurements are made at the lowest permissible analyte concentration and one measurement is made at the highest permissible analyte concentration. If the most accurate estimate of the slope is desired, then the measurements should be equally divided between the highest and lowest permissible concentration levels.

Agterdenbos (8) considered the effect of altering the concentrations of calibration samples on the precision obtained in the final concentration estimate. Both the distribution of the calibration measurements over the concentration range of interest and the number of replicate measurements were found to influence the results obtained in the subsequent quantitation step. A new quantity, the eccentricity, can be defined to describe the relationship between the precision of the estimated sample concentration and locations of the calibration points. The precision of the estimated sample concentration, Δx, is a function of several parameters: the selected statistical significance level, t; the standard deviation of the analytical procedure, s; the total number of calibration measurements, N; the number of replicate measurements of the sample, n; and the location of the sample measurement in the calibration range or the eccentricity, E.

$$\Delta x = 2\ t\ s\ (N^{-1} + n^{-1} + E)^{1/2} \qquad (3)$$

The eccentricity, E, can be calculated from the following relationship:

$$E = (\hat{\bar{x}}_s - \bar{x})^2 \; / \sum_{i=1}^{N} (x_i - \bar{x})^2 \qquad (4)$$

where $\hat{\bar{x}}_s$ is the mean estimated analyte concentration, and $\bar{x}$ is the mean concentration of the calibration samples. This is equivalent to the center of gravity of the calibration plot. From these relationships it is clear that the minimum uncertainity in the estimated analyte concentration will occur when the sample concentration is equal to the mean concentration of the calibration samples, in which case the eccentricity is equal to zero. As the estimated sample concentration moves to a value further from the center of gravity of the calibration plot the eccentricity increases and the precision of the estimated sample concentration becomes poorer.

In many analytical procedures, the assumption of homogeneity of variances may be false. For example, the precision obtained in spectrophotometry may be limited by the measurement readout error, detector shot noise, or source flicker noise (9). The classical approach to estimating measurement precision has assumed that the readout error of a linear transmittance scale is the dominant factor; in modern instruments it is far more likely that the dominant factor will be the photomultiplier shot noise.

Agterdenbos (6) has suggested that chemists give more care to the proper selection of the calibration relationship being used when performing a least squares analysis. One method for obtaining the calibration line when the precision of the response measurement is dependent on the analyte concentration is to use a weighted least squares procedure. Weighted least squares regression is analogous to ordinary least squares. Both methods are based on minimizing the sum of the squared deviations between the actual

responses and the calibration line. However, when weighted least squares is used, each residual is multiplied by a weighting factor, w_i, proportional to the reciprocal of the variance of the corresponding response measurement, r_i. The relationship to be minimized is now given as

$$D^2 = \sum_{i=1}^{N} w_i \, (r_i - \beta_0 c_i - \beta_1)^2 \qquad (5)$$

Schwartz (10) has illustrated the potential for nonuniform variance in both spectroscopic and chromatographic experiments. He concluded that if the analyst ignores variance nonuniformity, roughly the same calibration curve will be obtained. However, the confidence bands around the estimated analyte concentration may be severely in error at the extremes of the calibration curve. Garden and co-workers (11) have shown how weighted least squares procedures can improve the precision of the estimated analyte concentration when compared to ordinary least squares calibration.

Deterministic Error

In addition to the statistical errors which may arise from the improper application of least squares methods, the single-component linear model may also be affected by various types of deterministic errors. In most cases these deterministic errors can be traced to violation of the initial assumptions underlying the original model. Often if the source of the error can be identified, the calibration model can be adjusted to bring into consideration the effects of these additional factors.

One serious problem which occurs frequently in analytical chemistry is the presence of a sample matrix effect. This can be defined as a difference in the sensitivity coefficient, k, between the unknown sample being analyzed and the calibration standards. The interaction of the analyte with the sample matrix results in a change in the slope of the calibration plot.

The method of standard addition is widely used in analytical chemistry to address this particular type of problem. The response of the unknown sample is measured, a known amount of the pure analyte is added to the sample, and then the response of the sample after this addition is measured. The initial response measurement, r_0, depends only on the unknown concentration, c_0. After the addition is made the response is a function of both the original analyte present and the amount added. The matrix-corrected sensitivity coefficient is obtained from the change in the response due to the addition of pure analyte. This method still requires the response to be linearly related to the concentration and the sensor to be totally specific for the analyte of interest.

Different groups of workers have applied statistical techniques to calculate the optimum method of making the additions and the resulting precision in the estimate of the analyte concentration (12,13). The optimal size of the addition to be made is a function of the precision of the sensor and the form of the calibration function. Franke, de Zeeuw, and Hakkert (14) concluded that if a single addition is made, then optimum precision is obtained by making an addition of the largest possible amount of standard without exceeding the linear range of the calibration curve and making an equal number of replicate measurements before and after the addition.

The single-component linear model assumes that the analytical sensor posesses total specificity for the analyte of interest. Implicit in this

assumption is a requirement that the response of the sensor at zero analyte concentration is zero response units, or simply stated, the sensor can be zeroed. Two types of problems may lead to failure of this assumption: first, an instrumental or constant background; and second, a sample- or volume-dependent background.

An instrumental background will result in the addition of a constant non-volume-dependent term, d, to the simple linear model.

$$r = k\,c + d \qquad (6)$$

In a spectophotometric analysis, this constant term may arise from the use of mismatched optical paths or cells, temperature differences between the sample and calibration solutions, amplifier offsets, or similar problems. An instrumental background will cause a bias in the concentration estimate obtained from either a normal calibration or a standard addition experiment. Fortunately, this type of background can be handled by standard dilution.

A significantly more difficult problem is the presence of a sample background. In this situation the sensor no longer posesses complete specificity, but responds both to the analyte of interest and also one or more other components present in the sample. This problem has given rise to a multitude of separation and purification techniques directed at eliminating potential interferences. If the identities of the additional components are known and standards of these components are available, then the situation can be treated as a multicomponent analysis problem. However, if the identities of any of the interferents are not known or if calibration with these components is not possible, then the situation represents a sample- or volume-dependent background problem.

Multi-Component Linear Model

The multicomponent linear model is simply a generalization of the familiar single-component model. The responses due to each of the components present in the sample are assumed to add linearly, or can be transformed to yield the total response for any analytical sensor. For the case of two sensors, which respond to both of two analytes, a system of two equations is obtained. This can be written as

$$r_1 = \sum_{i=1}^{2} c_i k_{i1} = c_1 k_{11} + c_2 k_{21}$$
$$r_2 = \sum_{i=1}^{2} c_i k_{i2} = c_1 k_{12} + c_2 k_{22} \qquad (7)$$

where r_j is the response measured at the j-th sensor, c_i is the concentration of the i-th analyte, and $k_{i,j}$ is the sensitivity coefficient of the j-th sensor for the i-th analyte. Each equation represents the measured response for a single analytical sensor as the sum of the responses due to the individual components. For a mixture of N components, this model is generally written in matrix form as follows

$$r' = c'K \qquad (8)$$

The vector r is a column vector containing the response of the sample measured with P different sensors. The vector c is a column vector containing the concentrations of the N analytes present in the sample. The prime denotes the transpose of a matrix or vector. The matrix K contains the sensitivity coefficients for the N components at each of the P analytical sensors. Each row of the K matrix contains the P sensitivity coefficients of a single analyte. Each column of the K matrix contains the sensitivity coefficients of all N components for the same analytical sensor.

Several assumptions are implicitly made when the multicomponent linear model is used. These assumptions are analogous to the assumptions made with the single-component linear model. First, the response of each sensor is assumed to be linearly related to all analyte concentrations over the concentration ranges of interest. Second, the response due to each component present in the mixture sample is independent of the other N-1 components. Third, the response of each sensor can be zeroed. Lastly, the sensitivity coefficients do not change between the calibration and quantitation steps.

Classification of Samples

Martens et.al. (15) proposed the classification scheme for multicomponent mixtures shown in Table I. Mixture samples in which the individual component concentrations are known will be designated as class 1 samples; those samples in which the component concentrations are not known will be designated as class 2 samples. Class 1 samples are further grouped into two types. The first group is class 1A. These samples are fully defined, both the individual component concentrations and the pure component sensitivities are known to the analyst. In the second group, class 1B, the individual component concentrations are known but the pure component sensitivities are unknown. If N components are present in the mixture, then estimation of the individual component sensitivities requires either N pure samples or N mixtures of known composition. This class of samples represents the general problem of multicomponent calibration.

Class 2 samples are also grouped into two types. The first group, class 2A, are samples in which the component concentrations are unknown but the individual pure component sensitivities are known. This type of sample is representive of a multicomponent quantitation problem. The second group, class 2B samples, are mixtures in which neither the individual component concentrations nor their sensitivities are known by the analyst.

Analysis of a single class 2B sample is not possible. However, if a set of class 2B samples are available in which the relative concentrations of the components varies from sample to sample, then the methods appropriate to the multicomponent bilinear model may be used to obtain regions of physically allowable pure component sensitivities and concentration. An unambiguous solution for the sensitivities and individual component concentrations is not possible unless the analytical chemist can obtain further information about the samples.

Calibration

Several different methods are available for calibration in a multicomponent analysis. Kaiser (16) has grouped calibration methods into three main approaches: 1) σ-calibration, or calibration with synthetic standards; 2) α-calibration, or calibration with analyzed standard samples; and 3) δ-calibration, or calibration by differential additions. Of these three methods, calibration with synthetic standard samples is in Kaiser's view the most fundamental. Given a mixture of N components whose response can be measured at each of P different analytical sensors, the question arises of selecting the best method for first performing the calibration and then estimating the N analyte concentrations. If samples of the N pure components are available, then the simplest method of obtaining the sensitivity coefficients is to individually measure the response of each pure compound. This method may not always be possible. In some cases the pure substances may be very difficult or expensive to obtain and purify or the mixture may include analytes which are unstable in purified form. If the pure analytes are not available, but it

is possible to obtain a set of pre-analyzed standard samples, then the calibration can be based on comparison to these standard mixtures. Finally, if matrix effects are known to be present, the most appropriate method is to use a standard addition analysis to allow calibration within the sample matrix.

Multicomponent calibration based on the analysis of a set of mixture samples of known analyte concentrations to obtain the calibration relationship is frequently used. If a well characterized set of M mixture samples is obtainable, the sensitivity coefficients can be obtained by ordinary or weighted least squares multiple regression. The normal representation of this problem is as follows

$$R = C\,K \qquad (9)$$

where R is an M x P matrix of measured responses for each of the M mixture samples, C is an M x N matrix containing the N analyte concentrations for each of the mixtures, and K is defined as before. Since the concentrations of all of the analytes in each mixture sample are known and the mixture responses can be measured, the sensitivity matrix, K, can be obtained by multiplying both sides of equation 9 by the inverse of C. If there are the same number of mixture samples as analytes present, i.e. M = N, then this system of linear equations is exactly determined and the calibration step requires inverting C to yield

$$K = C^{-1}R \qquad (10)$$

However, if the number of calibration mixtures used is greater than the number of analytes, i.e. M > N, then the best estimate of the sensitivity matrix, K, is generally calculated from least squares multiple regression in

matrix form. The generalized inverse solution for the sensitivity matrix is given as

$$K = (C'C)^{-1}C'R \qquad (11)$$

In order to obtain the sensitivity matrix, K, from either equation 10 or 11, the number of analytical sensors, P, must be greater than or equal to the number of analytes. The analyte concentrations in a unknown mixture sample can be obtained by measuring its response and multiplying the transposed response vector by the generalized inverse of K,

$$c' = r'K'(KK')^{-1} \qquad (12)$$

For the entire analysis, calibration and quantitation of N analytes, this procedure requires at least N mixture samples, and inversion of two N x N matrices; C'C and KK'.

Brown and associates (17,18) have proposed an alternative formulation of the matrix multicomponent model, where instead of considering the response as a function of concentration, they consider the concentration a function of the measured response. This is written as

$$C = R P \qquad (13)$$

where the matrices C and R are defined as before and the matrix P represents the proportionality between C and R. The matrix P will have dimensions of P x N, i.e. sensors by analytes. For this model, the calibration step is expressed as

$$P = (R'R)^{-1}R'C \qquad (14)$$

which requires the inversion of the P x P matrix, R'R. Quantitation of an unknown sample is accomplished directly by multiplying the response vector, r, by the calibration matrix, P, to yield

$$c' = r'P \qquad (15)$$

The authors state that this method has the advantage of requiring only one matrix inversion instead of the two required by the conventional notation. Subsequently, they used this method for the spectrophotometric analysis of serum lipids with 85 calibration samples and 15 analytical wavelengths (19).

The difficulity in this analysis lies in the relative dimensions of the various matrices. In order to obtain a solution of equation 14 there must be more calibration mixtures than sensors being used. This is a drawback when the availability of diode array spectrophotometers makes it possible to measure 256 or more wavelengths as easily as four or five. In order to use all of the available wavelengths, one calibration sample must be prepared for every sensor used. Additionally, as the number of calibration samples and wavelengths are increased the size of the matrix R'R, which must be inverted in the calibration step, is also increased. If the sensors themselves are highly correlated or if the number of analytes is much less than the number of sensors, then the matrix R'R may have an effective rank of much less than P. In this situation, R'R will be almost singular and the inversion will be numerically unstable.

The method of principal component regression can be used as an alternative to ordinary least squares multiple regression (20). This method is based on replacing the M x P response matrix, R, with the product of two smaller matrices, T and B. Equation 13 can now be written as

$$C = T B P \qquad (16)$$

where the matrix T has dimensions M x A and the matrix B has dimensions A x P, with A << N and A << P. Decomposition of R into the matrices T and B is called singular value decomposition, eigenvector projection, factor analysis, or principal component analysis, depending on the scaling of R. The matrices T and B are selected in order to represent R as closely as possible and such that the columns of T and the rows of B are both orthogonal. Geometrically, the decomposition of R into T-B can be considered as a projection of the original data points, or mixture spectra, from a P-dimensional measurement space into a smaller A-dimensional space. The matrix T, whose elements are sometimes called the factor scores, contains the coordinates of the data points in the new A-dimensional space and the matrix B, containing the factor loadings, is the rotation matrix used to perform the projection. Solution of the original calibration problem now requires the inversion of T'T instead of R'R. Since the columns of T are orthogonal, this inversion is numerically well conditioned. This yields a calibration matrix G instead of the calibration matrix P.

$$G = B\,P = (T'T)^{-1}T'C \qquad (17)$$

The desired calibration matrix P can then easily be found as

$$P = B'G \qquad (18)$$

The quantitation step is exactly the same as used by Brown and co-workers (17,18).

A different approach to the multicomponent calibration problem, called partial least squares in latent variables (PLS), has been suggested by S. Wold and co-workers (21). PLS was developed by H. Wold (22) to solve complex data analysis problems in econometrics and psychometrics. It is somewhat analogous to principal component multiple regression in that the independent

variables, in this case the matrix R, are described by a principal component type model and then combined with a regression relationship relating the responses to the analyte concentrations contained in the matrix C. The difference is that in the PLS method the projection T is computed not only to model R but also to maximize its correlation with C. In principal component regression, T is selected only to model R. The PLS method involves first scaling both the response matrix, R, and the concentration matrix, C, such that the standard deviation of each column in these matrices is equal to one. The matrices are then centered by subtracting the average for each column. Each matrix is then modeled as a linear combination of new orthogonal latent variables. The latent variables are calculated by an iterative method which does not involve an explicit regression step. The maximun number of latent variables is the actual number of independent variables; however, normally fewer latent variables are used to allow filtering of the noise present in the data set. The PLS model is described as follows

$$c_{ij} = \bar{c}_j + \sum_{l=1}^{A} u_{il} b_{lj} + \epsilon'_{ij} \qquad (19)$$

$$r_{ik} = \bar{r}_k + \sum_{l=1}^{A} t_{il} d_{lk} + \epsilon''_{ik} \qquad (20)$$

$$u_{il} = \rho_l t_{il} \quad \text{for all } l=1,\ldots A \qquad (21)$$

where $u_{i,l}$ and $t_{i,l}$ are the latent variables and $b_{i,j}$ and $d_{i,k}$ are the loadings used to describe the concentration and response matrices, respectively. Equations 19 and 20, known as the outer relationship, describe the projection of the original variables into an A-dimensional space. Equation 21, known as the inner relationship, describes the correlation between the latent variables. The quantitation step in PLS is accomplished by first centering and scaling the measured response spectrum of an unknown mixture, calculating the latent variables, $t_{i,l}$, from the

loadings, $b_{i,j}$, calculating the latent variables, $u_{i,l}$, from the inner relationship, and then estimating the concentrations from equation 19. The PLS method has been compared to principal component regression for the multicomponent calibration and quantitation of spectrofluorimetric data from mixtures of humic acid and ligninsulfonate by Lindberg and co-workers (23). They concluded that: first, PLS was computationally faster than principal component regression; second, PLS calibrations have better predictive qualities since the method extracts information which has predictive relevance for the concentrations of the calibration mixture; third, a criterion could be established for determining if the calibration model was appropriate for a given unknown mixture; and fourth, like other methods based on principal component analysis, PLS was able to compensate for unidentified fluorescent species in the solution. This final conclusion implies that an analyte can be quantitated in the presence of a totally unknown background, but the experimental data reported does not support this conclusion.

It was already noted that matrix effects can affect the accuracy of the calibration in a single component analysis. Exactly the same difficulties may arise with the multicomponent linear model. As Kaiser (16) noted, standard additions provide the most appropriate calibration method if matrix effects occur. When discussing the single component linear model, it was observed that in most cases, the well known standard addition method was able to correct for these matrix effects, but the simple standard addition method required a fully selective sensor. Saxberg and Kowalski (24) have developed a multicomponent extension of the standard addition method which they named the generalized standard additon method or GSAM. The generalized standard addition method has two distinct

advantages: first, it allows the use of non-selective analytical sensors; and second, it corrects for the presence of matrix effects. The first advantage is a byproduct of the multicomponent nature of the method. This does not require the individual sensors to be fully selective for any one analyte, however it does require that the sensors do not respond to components in the sample of which additions have not been made. The second advantage is the result of using standard additions and making all the measurements within the sample matrix in order to obtain the sensitivity matrix, K. The response of each sensor is normally assumed to obey the linear multicomponent model given in equation 8, but models involving higher dimension polynomial relationships between the concentration and absorbance were described.

Experimentally, GSAM requires that M additions are made to the sample being analyzed. Each addition may contain one or more of the pure analytes, however, the additions must be made such that each pure analyte is added to the sample at least once. After each addition the response at each of P analytical sensors is measured. The response after the m-th addition is modeled as

$$r_m' = c_m' K \qquad (22)$$

where r_m is a column vector containing the measured responses, and c_m is a column vector containing the total concentrations of analyte present (c_0 + Δc). The response matrix, R, and the concentration matrix, C, are defined as

$$R' = [r_1, r_2, r_3, \ldots r_M] \qquad (23)$$

$$C' = [c_1, c_2, c_3, \ldots c_M] \qquad (24)$$

This allows a simple formulation of the problem as

$$R = C\,K \tag{25}$$

Each row of R and C corresponds to a separate multiple standard addition. The matrix R is always known. The matrix C is unknown since each row includes the unknown analyte concentration plus the amount of analyte which has been added. The matrix of sensitivity coefficients, K, is also unknown. Solution of this linear multiple linear system is accomplished by separating the terms as follows

$$C = \Delta C + C_0 \tag{26}$$

$$R = \Delta R + R_0 \tag{27}$$

where C_0 and R_0 are matrices with all rows identical to the initial concentration and initial response vectors c_0 and r_0, respectively, and ΔC and ΔR are the matrices of the net change in concentrations and responses due to the standard additions. ΔC and ΔR are always known to the analyst, hence the sensitivity coefficients can be calculated from

$$\Delta R = \Delta C\,K \tag{28}$$

The calibration step in GSAM is equivalent to the solution of this linear system. Assuming $N \neq M$, the solution is found by

$$K = (\Delta C'\Delta C)^{-1}\Delta C'\Delta R \tag{29}$$

or if $N = M$, then ΔC can be inverted directly. The quantitation step is given by equation 12. This is identical to the earlier discussion of the least squares matrix solution of the multicomponent model. It must also be noted that the matrix ΔC contains the effective concentration changes after

each standard addition. Unless the volume changes are negligible, ΔC cannot be known since the initial concentrations are not known. This problem can be avoided by incorporating a simple volume correction into the GSAM to convert from analyte concentrations to absolute quantities. Equation 22 is now written as

$$r_m' = (1/V_m)n_m'K \qquad (30)$$

where the vector n_m contains the absolute quantities, in grams or moles, of each analyte in a volume, V_m. Multiplying both sides of this equation by V_m leads to

$$q_m' = V_m r_m' = n_m'K \qquad (31)$$

where the vector q_m contains the P volume corrected responses. The remaining equations are obtained by substituting the volume corrected responses; q, Q, and ΔQ, for the responses; r, R, and ΔR, and by substituting the absolute quantities; n, N, and ΔN, for the concentrations; c, C, and ΔC.

Several more recent papers have examined the error propagation and statistical aspects of using the GSAM. Jochum, Jochum, and Kowalski (25) have stated the accuracy of GSAM in obtaining valid estimates of the initial analyte concentrations is dependent on at least five distinct factors: first, the accuracy of the response measurements; second, the accuracy and precision of the multiple standard additions; third, the magnitude of the interanalyte response interferences; fourth, the experimental design; and finally, the mathematical algorithims used in the computations. The first two of these factors are no different than the considerations required for any analytical method. The final factor, selection of the mathematical algorithims, can affect the results by introducing round-off errors into the computations.

The upper bound on relative errors in the estimated concentrations was found to depend on the condition number of both the calibration matrix, K, and the experimental design, which is described by the addition matrix, ΔN. The condition number of any nonsingular matrix A is defined as

$$\text{cond}(A) = ||A||\ ||A^{-1}|| \qquad (32)$$

where ||A|| is the Euclidian norm of the matrix A. If the matrix A is rectangular, then its condition number is given as

$$\text{cond}(A) = [\text{cond}(A'A)]^{1/2} \qquad (33)$$

It is important to note that the condition number of any matrix is always equal to or greater than one. In the GSAM experiment, the K matrix is determined by the solution of an overdetermined system of linear equations and therefore this matrix is not exactly known. Jochum, Jochum, and Kowalski showed that errors in the response measurements can be amplified by the chemist's choice of experimental design. An estimate of the error in the calculated K matrix was found to be

$$\frac{||\bar{k}-k||}{||k||} < \text{cond}(\Delta N)\ \frac{||\overline{\Delta q}_1 - \Delta q_1||}{||\Delta q_1||} \qquad (34)$$

where Δq_1 and $\Delta \bar{q}_1$ are the projections of Δq and $\Delta \bar{q}$ onto the range of N. A modification of the computational algorithim, called the incremental difference calculation, was described which minimized the error amplification due to the experimental design. In the incremental difference calculation the ΔQ matrix is composed of the change in volume corrected response between two successive additions and the ΔN matrix is composed of the absolute quantity of analyte added in a single addition. After scaling the condition number of the ΔN matrix is equal to one, which results in no error amplification being introduced in the final concentration estimates due to the

experimental design.

The condition number of the K matrix can also lead to a magnification of the potential concentration errors. The authors showed that, in the worst case, a small relative error in the initial response vector, r_0, could be magnified by the cond(K) to produce a larger relative concentration error. The error in the concentration estimates was found to be

$$\frac{||\delta c_0||}{||c_0||} < \text{cond}(K) \left[\frac{||\delta r_0||}{||r_0||} + \frac{||\delta k||}{||k||} \right] \qquad (35)$$

where δc_0, δr_0, and δk are the errors present in c_0, r_0, and K, respectively. Recently, Kalivas (26) showed the condition number of the K matrix is a extremely useful tool for assessing the analytical cost in terms of relative uncertainty of varying sensor selectivity. Minimization of the condition number of the K matrix can be used as a criteria for the selection of the optimal set of sensors for a particular multicomponent analysis.

Moran and Kowalski (27) have considered the statistical aspects of the GSAM. They have found that the uncertainty in the estimates of the sensitivity coefficients, i.e. the $k_{i,j}$'s, is dependent on three terms; the measurement variance, correlation of the response measurements due to subtraction of the initial response, and variance arising from the volume increase as a result of making standard additions. In order to reduce the variance and obtain the best possible accuracy in the concentration estimates, they recommend several steps. First, the volume increases must be minimized. Second, if random noise is the dominant source of error, then the total difference calculation method should be used. Third, the largest possible additions of analyte should be made.

Quantitation

Presuming the sensitivity matrix, K, has been obtained, the quantitation step can be approached by an extension of the single component model. Sternberg, Stillo, and Schwendeman (28) have described the application of the least squares method in matrix form to the spectrophotometric analysis of a five component mixture. They noted certain restrictions are necessary to assure a solution to the matrix problem will exist. The length of the response vector, r, and the column dimension of the sensitivity matrix, K, must be equal to or greater than the number of analytes, therefore P must be greater than or equal to N. In addition the rank of the sensitivity matrix must be at least N, which implies the P sensors must span a minimum of an N-dimensional space. If there are exactly the same number of sensors as there are analytes present, i.e. N = P, then the solution to the matrix problem is simply given as,

$$c' = r' K^{-1} \qquad (36)$$

However, if more sensors than the minimum number necessary to obtain a solution to the system of linear equations are used, i.e. P > N, then the method of least squares can be used to obtain the set of estimated analyte concentrations which minimizes the difference between the measured responses and the responses predicted by the multicomponent linear model. The solution to this least squares problem in matrix form was given in equation 12. Two years later, Zscheile and co-workers (29) used the matrix form of the least squares method to examine a four component spectrophotometric system. In analyzing a system of RNA-constituents, they observed the stability of the concentration estimates was very dependent on the wavelengths selected for the analysis. The poor stability obtained with some sets of wavelengths was

attributed to linear dependence of the underlying pure analyte spectra. The best results were obtained when all the available wavelengths were used.

The same considerations regarding homogeneity of variance, which were necessary for the single component linear model, must also be made when the multicomponent model is used. Haaland and Easterling (30) applied a linear additive multicomponent model to the analysis of infrared spectra of xylene isomer mixtures. They observed the noise characteristics of most infrared detectors were such that the noise was generally constant and independent of the signal level. The signal measured by these detectors is in transmittance, which is then converted to absorbance. Since Beer's law is generally obeyed in this spectral region the absorbance is directly proportional to concentration, however, the precision of the absorbance measurements are not independent of the measured responses. To account for this non-homogeneity, Haaland and Easterling used a weighted least squares procedure. Expanding the absorbance signal as a Taylor series about the transmittance and retaining only the first two terms, they found the variance of the noise was proportional to the inverse of the square of the transmittance. Therefore, they performed the analysis by first weighting each measured response in the spectrum by a factor equal to its transmittance squared. The matrix form of the weighted least squares estimate of the analyte concentrations is given by,

$$c' = r'V^{-1}K'(K\ V\ K')^{-1} \qquad (37)$$

where the matrix V is a diagonal matrix containing the reciprocal of the weights. This method of weighting assumes the errors in the responses are independent but with different variances. If the response measurements are correlated, equation 37 may still be used, however, the matrix V is no longer diagonal (4).

Deterministic Errors

As was observed with the single component model, various types of deterministic errors may affect the multicomponent linear model. These errors, which may be due to chemical, e.g. matrix effects or interferences, or instrumental factors, e.g. drifting or non-zeroed sensors, result in violating the assumptions present in the linear additive response model. In two recent papers (31,32), Kalivas and Kowalski have extended the GSAM model to add one or more terms to the basic model which allow for the detection and correction of sensor drift occuring during the course of the analysis. The GSAM model with the inclusion of terms for so-called time additions is

$$r_{ml} = \sum_{j=1}^{N} c_{mj} k_{jl} + \sum_{i=1}^{W} t^i k_{N+1,l} \qquad (38)$$

where W is the polynomial order of the drift model. Volume correction was performed as earlier described. The N+1-st to W-th rows of the K matrix represent the coefficients of the drift model. The drift coefficients can be examined statistically in order to detect the presence of a drifting analytical sensor. Estimation of the initial analyte quantities is accomplished as before, after deletion from the K matrix of the rows containing the sensitivity coefficients for the time additions. Implementation of the drift correcting GSAM model requires augmenting the ΔN matrix with W rows containing the time elapsed since the initial response measurement raised to the appropriate power. This was accomplished by developing a completely automated system for making the standard additions, measuring the responses, and recording the elapsed time (32). In addition to implementing the time additions and drift correction, this system was designed to make the standard additions by weight instead of by volume in order to minimize the relative errors in measuring the amount of analyte added.

Implicit in the multicomponent linear model is an assumption that the response of the analytical sensors can be zeroed. Two types of model failure have been identified in connection with this assumption: first, an instrumental or constant background; and second, a sample or volume dependent background. Altering the multicomponent model to compensate for an instrumental background can be accomplished by adding a constant term for each sensor,

$$r' = c'K + d' \qquad (39)$$

where the vector **d** contains the background contribution at each of the P analytical sensors. Vandeginste et al. (33) has shown a dilution procedure can be used to correct for a constant background response. Equation 39 is rewritten for volume corrected responses as

$$V_0 r_0' = V_0(c_0'K + d) = n_0'K + V_0 d' \qquad (40)$$

where V_0 is the initial volume of the sample mixture, r_0 is the initial response vector, c_0 is the initial concentration vector, and n_0 is the vector of initial analyte quantities. A standard dilution is performed by adding a volume, Δv, of pure solvent to the mixture sample. Equation 39 can again be rewritten in terms of the volume corrected responses; however, now the total sample volume is $V_0 + \Delta v$. Since the absolute quantities of analyte present have not been affected by the dilution, the difference in the volume corrected responses, Δq, is simply

$$\Delta q' = q' - q_0' = \Delta v d' \qquad (42)$$

This relationship allows estimation of the constant background vector, **d**, since it is a function of only the added volume, Δv, and the vector of measured changes in the volume corrected responses, Δq.

The presence of additional components in the sample mixture gives rise to a sample or volume dependent background. This can be incorporated into the standard multicomponent linear model by adding additional terms which express the response as a function of the known analytes and the additional interferents. The expanded model is

$$r_l = \sum_{i=1}^{N} c_i k_{il} + \sum_{j=1}^{T} c_j k_{jl} \qquad \text{for all } l=1,\ldots P \qquad (43)$$

where r_l is ths response of the l-th sensor. The first summation, which runs from one to N, accounts for the response caused by the N known analytes. The second summation, which runs from one to T, accounts for the response caused by the presence of the T interfering components. Since the identities of the T interfering components are not known, no standards for these components can be used nor can standard additions of these components be made. Hence, during the course of the analysis, the relative amounts of the interferents with respect to each other will not change. Therefore, these T interferents can be replaced by a single term which represents their combined influence on the measured sensor responses,

$$r_l = \sum_{i=1}^{N} c_i k_{il} + f_l \qquad \text{for all } l=1,\ldots P \qquad (44)$$

where f_l is the combined background response at sensor l. The important distinction between this model and the model, given in equation 39 which describes an instrumental background, is that the term f_l is a function of the sample volume, therefore the standard dilution method used by Vandeginste does not apply. Since the sample background, f_l, is not known, an iterative method must be used to perform mixture quantitation.

In their original paper describing the GSAM model, Saxberg and Kowalski (24) discussed the problem of analytical sensors which were not zeroed. They observed that if the background response was small relative to the initial unknown response and the problem was reasonably insensitive to perturbations, then the effect on the final solution can be expected to be small. Leggett(34) has applied non-negative least squares regression and simplex optimization to multicomponent spectrophotometric data. He concluded either of these methods avoid the problem of negative molar absorptivities or concentrations which are sometimes obtained when ordinary least squares regression has been used. This conclusion was reached with the stated assumption that the correct model, e.g. all components were known, had been used to set up the analysis. Gayle and Bennett (35) carried out simulation studies to determine the effect of model departure in multicomponent analysis on the concentration estimates obtained by ordinary least squares regression, non-negative least squares regression, and linear programming. They observed that when various types of model failure were simulated, all three methods yielded biased results, with no single method being consistently superior to the other two. In addition, of the three methods attempted only ordinary least squares provided any indication that the model was not valid. Omission of significant terms in this model frequently led to negative analyte concentrations, a result which obviously had no chemical meaning. However, non-negative least squares regression and linear programming yielded results which at least on the surface seemed chemically plausable, but were also significantly in error.

The final type of model failure which may occur is a failure of the assumed linear relationship between analyte concentration and the measured

response. Apparent deviations from the ideal behavior described by the Beer-Lambert law, which is widely used in spectrophotometric analysis, are well known. Saxberg and Kowalski (24) developed the original GSAM model to allow the response to be either a linear, quadratic, or higher degree polynomial function of the analyte concentration. Unfortunately, as the number of terms in the model increases, so does the required number of standard additions and measurements which the analyst must make. An alternative approach has been used by Vandeginste and co-workers (33) involving the application of a mathematical technique known as Kalman filtering to provide continuous testing of the validity of the linear model during the data acquisition stage of a GSAM experiment. Poulisse (36) has also applied the Kalman filter to the analysis of multicomponent spectrophotometric mixtures. Seelig and Blount (37,38) have applied this method to anodic stripping voltammetry and S. Brown and co-workers (39,40) have used the Kalman filter with linear sweep voltammetry and photoacoustic spectroscopy. This filter relies on a recursive algorithim which constantly updates the estimated sensitivities as more standard mixtures are analyzed. The recursive nature of this filter, which has only recently seen application in analytical chemistry, has the potential of providing feedback for on line evaluation and optimization of the calibration process.

Multicomponent Bilinear Models

The multicomponent bilinear model is obtained when a second measurement dimension is incorporated into the multicomponent linear model. This model describes the response of a single mixture sample along two independent measurement axes. Important applications of the bilinear model in analytical

chemistry include instrumental techniques based on two spectroscopic measurements, e.g. fluorescence emission-excitation matrices or EEMs, and techniques based on a combination of chromatographic and spectroscopic measurements, e.g. LC/UV, GC/MS, or GC/FTIR analyses. The response of a single component can be described as

$$\mathbf{M} = \alpha \times \mathbf{y}' \qquad (45)$$

where **M** contains the measured responses and is the outer product of the vectors **x** and **y** multiplied by a concentration dependent factor, α. The vector **x** represents the spectral, chromatographic, or temporal profile in the first dimension and **y** represents the spectral, chromatographic, or temporal profile of the compound in the second dimension. For example, a GC/MS peak consisting of 50 mass spectra each composed of 20 distinct m/e ratios would result in a matrix of spectral intensities, **M**, containing 50 rows and 20 columns. The vector **x** would have 50 elements and describe the gas chromatographic elution profile. The vector **y** would have 20 elements and represent the mass spectrum of the pure compound. Normally, **x** and **y** are normalized to a length of one, so that the factor α is then proportional to the standard concentration of the pure compound. Assuming that each component in a N component mixture responds independently of the remaining N-1 components, the response of the mixture can be represented by

$$\mathbf{M} = \sum_{i=1}^{N} c_i \mathbf{M}_i = \sum_{i=1}^{N} c_i (\alpha \mathbf{x}\mathbf{y}')_i \qquad (46)$$

where $\mathbf{M}_i$ is the standard response matrix due to component i in the mixture and c_i is the concentration of the i-th component divided by its standard concentration.

Least Squares Multiple Regression

The analysis of the data obtained from an experimental system, which is described by the multicomponent bilinear model, depends on the information available to the analyst. A common problem is the quantitation of several components whose identities are known and whose standard matrices are available. In this situation, least squares regression may be used. The objective is to minimize the sum of the squared elements of the residual matrix, E, which is defined as

$$E_{ij} = M_{ij} - \sum_{k=1}^{N} \beta_k (M_{ij})_k \qquad (47)$$

where the parameters β_k are the amounts of each of the k compounds present in the mixture. Warner et al. has applied this method to the analysis of fluorescence emission-excitation matrices (41). The least squares approach, while easy to implement and conceptually simple, yields accurate results only if standards of all of the mixture components are included in the data analysis.

Rank Annihilation

In many situations, the identities of all components contributing to the measured response may not be known. Ho and coworkers have developed the method of rank annihilation to allow the quantitation of one or several components without requiring knowledge of all of the components in a mixture sample (42,43). They have applied this method to the quantitative analysis of multicomponent emission-excitation matrices (EEMs) obtained from the analysis of polynuclear aromatic hydrocarbon mixtures with the video fluorometer. Ideally the mixture matrix, M, should have a rank equal to the number of components, N, in the mixture. For a mixture of N components, the best least squares approximation of M is given by

$$M = \sum_{k=1}^{N} \xi_k u_k v_k' \tag{48}$$

where

$$Mv_k = \xi_k u_k \tag{49}$$

and

$$M'u_k = \xi_k v_k \tag{50}$$

The eigenvectors $\{u_1, \ldots u_N\}$ and $\{v_1, \ldots v_N\}$ should span the same vector spaces as the pure component vectors $\{x_1, \ldots x_N\}$ and $\{y_1, \ldots y_N\}$. The number of nonzero eignevalues, ξ_k, equals the number of components in the sample. In order to perform rank annihilation an amount, β, of the standard matrix, M_1, which corresponds to a component known to be present in the mixture, is subtracted from the mixture matrix, M, to yield

$$E = M - \beta M_1 \tag{51}$$

When the correct value of β, corresponding to the concentration of M_1 in M, has been subtracted the rank of EE' will be N-1. This is indicated by one of the nonzero eigenvalues in EE' approaching zero. Since real data contains experimental error, the eigenvalue does not become exactly zero, but it does have a distinct minimum. The advantages of this technique are that it does not require the knowledge of all of the sample constituents or the presence of selective spectral regions.

If quantitation of several known species in a multicomponent bilinear mixture are desired, then an extension of rank annihilation based on the Fletcher-Powell algorithm may be used (44). This algorithm allows simultaneous computation of the concentrations of all known components in the

mixture sample. McCue and Malinowski have used rank annihilation of UV absorbance spectra to quantify coeluting liquid chromatographic peaks (45). Applying rank annihilation to LC/UV data requires that the elution profiles of each individual component in the mixture are exactly reproducible between the chromatographed standard samples and the mixture.

Self Modeling Curve Resolution

In 1971, Lawton and Sylvester (46) reported a method, which they termed self modeling curve resolution, for resolving two unknown overlapping functions from an observed set of mixtures of the two functions. They noted that this type of problem arises frequently in areas such as chromatography and spectrophotometry. This method is based on the assumption that neither the identities of the individual components nor their responses are known, but the responses for a number of mixtures of varing relative amounts of the same underlying components have been measured. The objectives of self modeling curve resolution are two-fold: first, to estimate the spectra of the underlying pure components; and second to quantify the amount of each pure component present in a given mixture. The model developed by Lawton and Sylvester can be described as follows. The measured response of a mixture of two pure components can be expressed as the sum of the responses of the individual components. This is simply the two component case of the multicomponent linear model developed in the last section and can be written

$$m = x_1 y_1 + x_2 y_2 \qquad (52)$$

where m represents a single mixture spectrum, x_1 and x_2 are the concentrations of the two pure components,and the vectors y_1 and y_2 are the spectra of the pure components. Normalization of the pure component spectra does not restrict the shape of the unknown spectra. The concentrations x_1 and x_2 are now defined relative to the concentration of analyte which produces an absorbance

spectrum of unit area. If N different mixture samples of these two components are measured, then the entire data set can be expressed in matrix form as

$$\mathbf{M} = \mathbf{X}\,\mathbf{Y} \qquad (53)$$

where $\mathbf{M}$ is a N x P matrix of measured responses, $\mathbf{X}$ is a N x 2 matrix of analyte concentrations, and $\mathbf{Y}$ is a 2 x P matrix of analyte spectra scaled to unit area. Since only two components are present, each observed mixture spectrum, e.g. each row of $\mathbf{M}$, can be expressed as a linear combination of the first two eigenvectors of the second moment matrix, $\mathbf{M'M}/N$. That is

$$\mathbf{m}_i = \epsilon_{i1}\mathbf{V}_1 + \epsilon_{i2}\mathbf{V}_2 \qquad (54)$$

where $\mathbf{m}_i$ is the i-th mixture spectrum and $\mathbf{V}_1$ and $\mathbf{V}_2$ are the eigenvectors associated with the two largest eigenvalues of $\mathbf{M'M}/N$. The spectra, $\mathbf{y}_1$ and $\mathbf{y}_2$, of the two pure components must also be linear combinations of these two eigenvectors.

$$\mathbf{y}_i = \eta_{i1}\mathbf{V}_1 + \eta_{i2}\mathbf{V}_2 \qquad \text{for } i=1,2 \qquad (55)$$

Determination of the values of η_{i1} and η_{i2} is equivalent to estimation of the unknown pure spectra.

Lawton and Sylvester applied three restrictions in order to obtain physically meaningful estimates of the pure spectra, $\mathbf{y}_1$ and $\mathbf{y}_2$. The first restriction was that all elements of the unknown pure spectra must be non-negative. This implies that η_{i1} and η_{i2} must satisfy

$$\eta_{i1}v_{1k} + \eta_{i2}v_{2k} \geq 0 \qquad \text{for all } k=1,\ldots P \qquad (56)$$

where v_{jk} is the k-th element of eigenvector $\mathbf{V}_j$. This is equivalent in a chemical sense to not allowing negative absorbances. The second restriction

was that all of the mixture spectra must be composed of non-negative amounts of the two pure components. This requires $x_{ij} > 0$ for all i and j. From equation 52, 54, and 55, it can be shown this restriction is equivalent to requiring $x_{i1} > 0$ and $x_{i2} > 0$ in

$$(\epsilon_{i1},\epsilon_{i2}) = x_{i1}(\eta_{11},\eta_{12}) + x_{i2}(\eta_{21},\eta_{22}) \text{ for all } i=1,\ldots S \quad (57)$$

The final restriction was based on the assumption that the unknown spectra, y_i, have been normalized to unit area. Figure 1 illustrates these three restrictions plotted in the 2-dimensional eigenvector space $\{V_1,V_2\}$. The angle formed by the inner constraint in figure 1 represents the range of relative analyte concentrations within the set of mixture samples. The angle formed by the outer constraint is related to the spectral uniqueness of the two pure components. Without requiring any assumptions as to the shape of the spectral curves, two regions, F_I and F_{II}, which contain the eigenvector representation of the pure spectra, y_1 and y_2, were obtained.

Sharaf and Kowalski (47,48) have considered the problem of quantitation in the two dimensional eigenvector space. They have shown that quantitative resolution of the two components in any given mixture spectrum is a straight foward function of the relative positions of the two pure spectra and the mixture spectrum in the eigenvector space. Assuming the mixture spectrum has been normalized to unit area, it will fall somewhere along the line segment separating regions F_I and F_{II} in figure 1. In order to quantify a mixture spectrum, the positions of the pure spectra, y_1 and y_2 within the regions F_I and F_{II}, respectively, must be known or estimated. If point m in region F_I is selected to represent the pure spectrum of component 1 and

point n in region F_{II} is selected to represent the pure spectrum of component 2; then Sharaf and Kowalski proved the fraction of the total response of mixture i due to component 1, F_{i1}, is given by

$$F_{i1} = d_{ni} \; / \; d_{mn} \qquad (58)$$

where d_{ni} is the euclidean distance from point n to mixture i and d_{mn} is the distance from point m in region F_I to point n in region F_{II}. The analogous expression for the fraction of the total response of mixture is due to component 2, F_{i2}, is given as

$$F_{i2} = d_{mi} \; / \; d_{mn} \qquad (59)$$

The major problem to be addressed in quantitating mixture spectra is the selection of the points m and n to be used as the best estimates of the pure spectra, y_1 and y_2. Sharaf and Kowalski considered several possibilities. First, if the width of the solution bands, F_I and F_{II}, are equal to zero, then pure spectra of both components have been measured and at least one specific sensor (e.g. wavelength, mass/charge ratio) exists for each component. In this case no assumptions are necessary to correctly quantify the mixture spectra. Second, if the solution band widths are not zero but specific sensors are known to exist, then all measured samples are mixtures of both components. Since specific sensors are known, the outer edges of the solution bands are the correct choice for the estimates of the pure component spectra. Third, if the solution band widths are not zero and specific sensors are not known to exist, then some assumptions must be made in order to quantify the mixture. The authors recommended using the inner edges of the solution bands, e.g. the purest spectra recorded, as an initial estimate of the pure component spectra. Alternately, the mid-points of each region may be used in the absence of further information.

A similar approach to curve resolution has been used by Martens (49). The major difference in the model used by Martens compared to that used by Lawton and Sylvester is that prior to extracting the eigenvectors of the moment matrix, Martens normalized the mixture spectra to constant area and centered the data matrix by subtracting the mean response of each sensor. This resulted in one less eigenvector being required to represent the mixture spectra in the reduced eigenvector space. Therefore, a mixture spectrum containing two underlying components can be represented by a linear combination of the mean and the first eigenvector of the centered covariance matrix. The advantage of this additional step is two-fold: first, one less dimension is necessary to represent the data, hence the factor analysis solution is somewhat easier to interpret; and second, the large trival variance associated with the mean has been removed by centering the data. Martens has made the same assumptions as were used by Lawton and Sylvester: first, only non-negative responses are allowed; second, only non-negative quantities of analytes may be present; and third, the pure component spectra are scaled to constant area. Spjøtvoll, Martens, and Volden (15) have compared the constraint equations for the two dimensional case using the mean plus one eigenvector model to the constraint equations as formulated by Lawton and Sylvester. When the mean plus one eigenvector model is used, quantitation can be accomplished using the method described by Sharaf and Kowalski (48). Osten and Kowalski (50) have recently examined the quantitative accuracy of self modeling curve resolution for the analysis of UV absorbance data obtained from a diode array high performance liquid chromatography detector.

Warner et al. (51) have used an approach similar to curve resolution which is based on the eigenanalysis of fluorescence emission-excitation

matrices. This method makes use of the same assumptions as the Lawton and Sylvester approach, only non-negative responses and non-negative quantities of each component are permitted. Warner and coworkers have relaxed these constraints allowing some elements to be slightly below zero in order to account for noise in the experimental data. Since the EEM represents data involving two spectral dimensions, they have considered the uncertainties in the estimated spectra for differing combinations of spectral overlap involving either one or both spectral dimensions between the two pure components.

The problem of generalizing curve resolution from the 2 component situations described above to the N component case is not trivial. Martens (49) examined the problem of three component mixtures of cereal amino acids. Ohta (52) has shown the solution of the 3-components problem for a mixture of photographic dyes. Very recently, Borgen and Kowalski (53) have described a general solution for the N-component resolution case. In all of these situations, the same non-negative quantity and non-negative response constraints have been utilized.

The multivariate methods discussed can be used to improve the precision and accuracy of an analytical procedure. The widespread incorporation of microprocessors in analaytical instrumentation can inundate the chemist with raw data. In order to obtain valid chemical information from this wealth of data, the analyst must consider not only the chemical system under evaluation but also the advantages, disadvantages, limits, and assumptions inherent in various potential data anlaysis approaches.

REFERENCES

1. Kowalski, B. R. Anal. Chem. 1980, **52**, 112R-122R.

2. Frank, I. E.; Kowalski, B. R. Anal. Chem. 1982, **54**, 232R-243R.

3. Kateman, G. Tr. in Anal. Chem. 1983, **3(2)**, IX-X.

4. Draper, N.; Smith, H. "Applied Regression Analysis", 2nd ed.; John Wiley and Sons: New York, 1981; chapter 1.

5. Nattrella, M. G. "Experimental Statistics", National Bureau of Standards Handbook 91; Government Printing Office, Washington, D.C., 1963; chapter 5.

6. Agterdenbos, J. Anal. Chim. Acta 1979, **108**, 315-323.

7. Shewell, C. T. Anal. Chem. 1960, **32**, 1535.

8. Agterdenbos, J. Anal. Chim. Acta 1981, **132**, 127-137.

9. Meehan, E. J. Treatise on Analytical Chemistry, Part 1 vol. 7, 2nd ed.; edited by Elving, P. J.; John Wiley and Sons: New York, 1981; section H, chapter 2.

10. Schwartz, L. M. Anal. Chem. 1979, **51**, 723-727.

11. Garden, J. S.; Mitchell, D. G.; Mills, W. N. Anal. Chem. 1980, **52**, 2310-2315.

12. Ratzlaff, K. L. Anal. Chem. 1979, **51**, 232-235.

13. Larsen, I. L.; Hartmann, N. A.; Wagner, J. J. Anal. Chem. 1973, **45**, 1511-1513.

14. Franke, J. P.; de Zeeuw, R. A.; Hakkert, R. Anal. Chem. 1978, **50**, 1374-1380.

15. Martens, H.; Spjøtvoll, E.; Volden, R. Technometrics 1982, **24**, 173-180.

16. Kaiser, H. Pure Appl. Chem. 1973, **34**, 35-61.

17. Brown, C. W.; Lynch, P. F.; Obremski, R. J.; Lavery, D. S. Anal. Chem. 1982, **54**, 1472-1479.

18. Maris, M. A.; Brown, C. W.; Lavery, D. S. Anal. Chem. 1983, **55**, 1694-1703.

19. Kisner, H. J.; Brown, C. W.; Kavarnos, G. J. Anal. Chem. 1983, **55**, 1703.1707.

20. Massy, W. F. J. Amer. Stat. Assoc. 1965, **60**, 234-256.

21. Sjostrom, M.; Wold, S.; Lindberg, W.; Persson, J.; Martens, H. Anal. Chim. Acta 1983, **150**, 61-70.

22. Joreskog, K. G.; Wold, H. Ed. "Systems Under Indirect Observation", Parts I and II; North Holland: Amsterdam 1982.

23. Lindberg, W.; Persson, J.; Wold, S. Anal. Chem. 1983, **55**, 643-648.

24. Saxberg, B. E. H.; Kowalski, B. R. Anal. Chem. 1979, **51**, 1031-1038.

25. Jochum, C.; Jochum, P.; Kowalski, B. R. Anal. Chem. 1981, **53**, 85-92.

26. Kalivas, J. H. Anal. Chem. 1983, **55**, 565-567.

27. Moran, M. G.; Kowalski, B. R. Anal. Chem. 1984, **56**, 562-569.

28. Sternberg, J. C.; Stillo, H. S.; Schwendeman, R. H. Anal. Chem. 1960, **32**, 84-90.

29. Zscheile, F. P.; Murray, H. C.; Baker, G. A.; Peddicord, R. G. Anal. Chem. 1962, **34**, 1776-1780.

30. Haaland, D. M.; Easterling, R. G. Appl. Spec. 1982, **36**, 665-673.

31. Kalivas, J. H.; Kowalski, B. R. Anal. Chem. 1982, **54**, 560-565.

32. Kalivas, J. H.; Kowalski, B. R. Anal. Chem. 1983, **55**, 532-535.

33. Vandeginste, B.; Klaessens, J.; Kateman, G. Anal. Chim. Acta 1983, **150**, 71-86.

34. Leggett, D. J. Anal. Chem. 1977, **49**, 276-281.

35. Gayle, J. B.; Bennett, H. D. Anal. Chem. 1978, **50**, 2085-2089.

36. Poulisse, H. N. J. Anal. Chem. Acta 1979, **112**, 361-374.

37. Seelig, P. F.; Blount, H. N. Anal. Chem. 1976, **48**, 252-258.

38. Seelig, P. F.; Blount, H. N. Anal. Chem. 1979, **51**, 327-337.

39. Brown, T. F.; Brown, S. D. Anal. Chem. 1981, **51**, 1410-1417. (correction) Anal. Chem. 1982, **54**, 607.

40. Rutan, S. C.; Brown, S. D. Anal. Chem. 1983, **55**, 1707-1710.

41. Warner, I. M.; Davidson, E. R.; Christian, G. D. Anal. Chem. 1977, **49**, 2155-2159.

42. Ho, C. N.; Christian, G. D.; Davidson, E. R. Anal. Chem. 1978, **50**, 1108-1113.

43. Ho, C. N.; Christian, G. D.; Davidson, E. R. Anal. Chem. 1980, **52**, 1071-1079.

44. Ho, C. N.; Christian, G. D.; Davidson, E. R. Anal. Chem. 1981, **53**, 92-98.

45. McCue, M.; Malinowski, E. R. J. Chrom. Sci. 1983, **21**, 229-234.

46. Lawton, W. H.; Sylvestre, E. A. Technometrics 1971, **13**, 617-633.

47. Sharaf, M. A.; Kowalski, B. R. Anal. Chem. 1981, **53**, 518-522.

48. Sharaf, M. A.; Kowalski, B. R. Anal. Chem. 1982, **54**, 1291-1296.

49. Martens, H. Anal. Chim. Acta 1979, **112**, 423-442.

50. Osten, D. W.; Kowalski, B. R. Anal. Chem. 1984, **56**, in press.

51. Warner, I. M.; Christian, G. D.; Davidson, E. R.; Callis, J. B. Anal. Chem. 1977, **49**, 564-573.

52. Ohta, N. Anal. Chem. 1973, **45**, 553-557.

53. Borgen, O. S.; Kowalski, B. R. manuscript in preparation.

CREDIT

This work was supported in part by the Office of Naval Research and by the National Science Foundation under Grant CHE-8004220.

Table I: Classification of Unconstrained Additive Mixtures

Class	Concentrations	Spectra
1A	known	known
1B	known	unknown
2A	unknown	known
2B	unknown	unknown

Figure 1. Generalized plot of two dimensional eignevector space. The outer edges of the shaded region represent the non-negative response constraint. The inner edges represent the non-negative quantity constraint. The regions F_I and F_{II} are the are the allowable regions for the location of the pure spectra m and n.

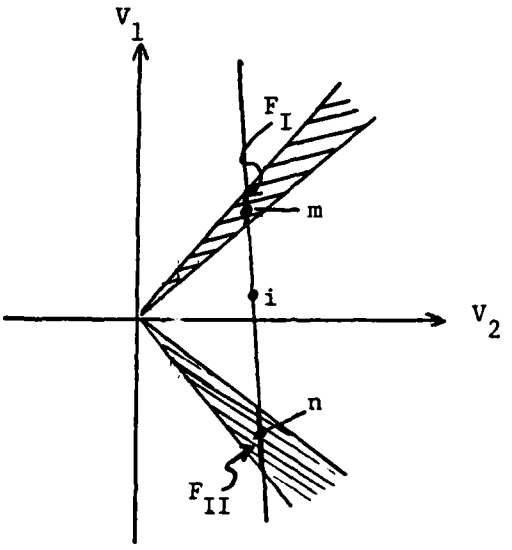


Figure 1.

DL/413/83/01
GEN/413-2

TECHNICAL REPORT DISTRIBUTION LIST, GEN

	No. Copies
Office of Naval Research Attn: Code 413 800 N. Quincy Street Arlington, Virginia 22217	2
Dr. Bernard Douda Naval Weapons Support Center Code 5042 Crane, Indiana 47522	1
Commander, Naval Air Systems Command Attn: Code 310C (H. Rosenwasser) Washington, D.C. 20360	1
Naval Civil Engineering Laboratory Attn: Dr. R. W. Drisko Port Hueneme, California 93401	1
Defense Technical Information Center Building 5, Cameron Station Alexandria, Virginia 22314	12
DTNSRDC Attn: Dr. G. Bosmajian Applied Chemistry Division Annapolis, Maryland 21401	1
Dr. William Tolles Superintendent Chemistry Division, Code 6100 Naval Research Laboratory Washington, D.C. 20375	1
Dr. David Young Code 334 NORDA NSTL, Mississippi 39529	1
Naval Weapons Center Attn: Dr. A. B. Amster Chemistry Division China Lake, California 93555	1
Scientific Advisor Commandant of the Marine Corps Code RD-1 Washington, D.C. 20380	1
U.S. Army Research Office Attn: CRD-AA-IP P.O. Box 12211 Research Triangle Park, NC 27709	1
Mr. John Boyle Materials Branch Naval Ship Engineering Center Philadelphia, Pennsylvania 19112	1
Naval Ocean Systems Center Attn: Dr. S. Yamamoto Marine Sciences Division San Diego, California 91232	1

END

FILMED

1-85

DTIC